STATISTIQUE INFERENTIELLE

Module Méthodes Quantitatives :

-Statistique III

(S4)

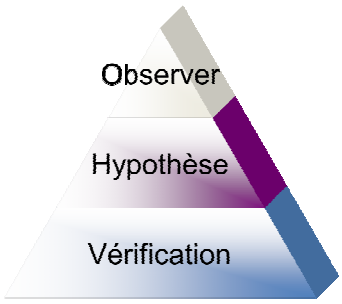
Section D

- On inverse les conclusions de l'étape précédente pour en déduire la structure vraisemblable de la population dont est issu l'échantillon observé; c'est la phase inférentielle (3ème chapitre)

INTRODUCTION GENERALE

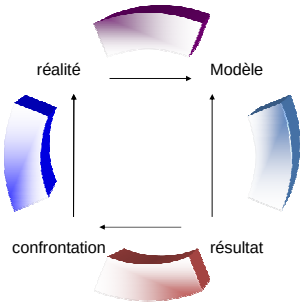
- Il y a en statistique deux approches :
- La Description (vue en S2) : consiste à résumer et à décrire un ensemble de données.
- L'inférence statistique : son but est d'étendre les propriétés de l'échantillon à la population entière (objet de ce semestre) et de valider ou de rejeter des hypothèses à priori ou formulées après une étape descriptive (objet de S5).

On remarque que cette démarche est semblable à la démarche scientifique habituelle



La démarche statistique est la suivante

- L'échantillon est tiré au hasard dans une population plus vaste (1er chapitre).
- Le calcul des probabilités (vu en S3) permet ensuite de préciser les caractéristiques de l'ensemble des échantillons que l'on aurait pu obtenir par le même procédé; c'est l'étude des distributions d'échantillonnage (2ème chapitre)



– Les méthodes statistiques sont aujourd'hui utilisées dans presque tous les domaines de l'activité humaine : économie, gestion, industrie, médecine, sciences humaines…

CHAPITRE 1

ECHANTILLONNAGE

PROGRAMME DE CE Semestre :S4

- **Chapitre 1 : L'échantillonnage**
- **Chapitre 2 : L'estimation**

Introduction

- L'enquête statistique est l'opération technique qui consiste à élaborer les statistiques. Elle a pour but de déterminer un ensemble de caractéristiques d'une population.
- On distingue deux types d'enquêtes:

Le recensement et *le sondage*

BIBLIOGRAPHIE

Titre	Auteurs	Code
Méthodes statistiques	B. Grais	stat22
Introduction à la statistique	J.P Bélisle ;J. Desrosiers	stat20
Théorie des sondages	C. Gouriéroux	stat49
Méthodes statistiques I	A. Vogt	stat25
Eléments de statistique d'aide à la décision	Hafidi et Touijar	

I- RECENSEMENTS ET SONDAGES

- 1- Définitions
  - **Définition1**: La population est un ensemble de personnes ou d'objets sur lesquelles porte une étude. Et on appelle individu chaque élément de cette population.
  - **Définition2**: On appelle recensement ou (enquête exhaustive) l'observation de la population entière.

– **Remarque**: Selon l'ONU: «le recensement de l'habitat est une opération qui permet de recueillir, grouper, évaluer, analyser et publier les données démographiques, économiques et sociales se rapportant à un moment donné à tous les habitants d'un pays »

– **Définition3**: On appelle *échantillon*, une partie représentative de la population observée.

– **Remarque**: La population d'où on tire l'échantillon s'appelle « population mère »

3- Types d'erreurs:

On distingue deux types d'erreurs:

a)- **Erreur de mesure (e_m):** elle provient des imprécisions du questionnaire, des erreurs professionnelles des enquêteurs…

b)- **Erreur d'échantillonnage** ou erreur aléatoire (e_a): elle tient au fait qu'on n'observe qu'une partie de la population.

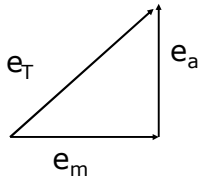
- Définition4: On appelle enquête par sondage, l'observation d'une partie représentative de la population mère, dans le but d'étudier un ensemble donnée de caractéristiques de cette dernière.
- Définirion5: On appelle taux de sondage, le rapport de la taille d'échantillon à la taille de la population mère:

$$\tau = \frac{n}{N}$$

n : échantillon

N : population

- **L'erreur totale** est la somme(vectorielle) des deux erreurs précédentes:



$$e_T^2 = e_m^2 + e_a^2$$

2- Les avantages des enquêtes par sondage

Les sondages présentent de nombreux avantages par rapport aux recensements. Leurs coût est nettement moins élevé. De plus, ils sont plus rapides. Seul le recensement exprime un résultat **certain** puisqu'il n' y a plus, en théorie, de problème d'inférence statistique (problème d'estimation).

Cependant, l'expérience montre que les sondages sont souvent très **précis.**

- **Remarque**: Dans un recensement, l'erreur aléatoire disparaît mais l'erreur de mesure persiste, et elle est souvent beaucoup plus importante que dans un sondage; car on doit employer un grand nombre d'enquêteurs hâtivement formés. Il n'y a donc aucune raison que l'erreur totale soit plus grande dans le cas d'un sondage.
- Un sondage bien fait peut-être plus précis qu'un recensement tout en coûtant beaucoup moins cher.

•Le principal inconvénient des sondages est de présenter des erreurs d'échantillonnage, qu'on tente de réduire en utilisant des méthodes rigoureuses de construction d'échantillons.

A- Echantillonnage Aléatoire

• **1- Echantillonnage aléatoire simple:**

– La construction d'un échantillon aléatoire simple de taille n est réalisée par un tirage au hasard avec remise de n individus dans l'ensemble de la population. Ainsi, tous les individus seront tirés de manière indépendante et auront une chance égale de faire partie de l'échantillon.

•Pour constituer un tel échantillon, on fait souvent appel aux « Tables des nombres aléatoires».

Quelques méthodes de prélèvement d'un échantillon

•On distingue deux types de sondages

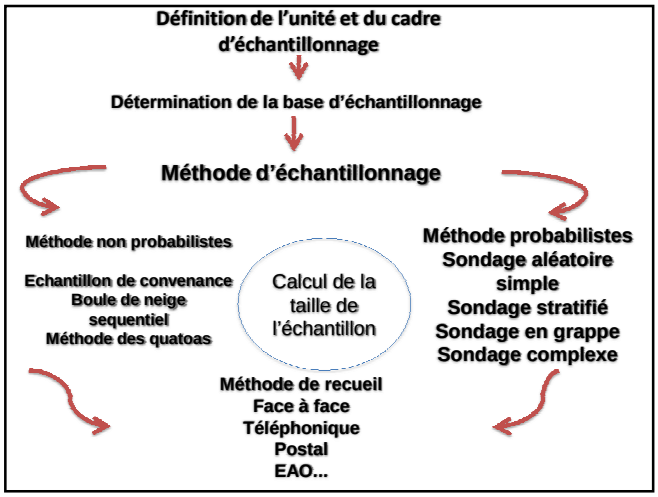
– **Les sondages aléatoires**: qui aboutissent à la construction d'échantillons aléatoires.

– **Les sondages par choix raisonné** génèrent des échantillons empiriques. Cette deuxième méthode ne sera pas étudiée en ce semestre.

B-Le tirage d'échantillon aléatoire simple

1- Les tirages avec et sans remise

– L'échantillonnage aléatoire simple (EAS) est basé sur un tirage avec remise. Si on constitue un tel échantillon, chaque individu aura une probabilité 1/N d'être le premier élément de l'échantillon. Pour le deuxième élément, chaque individu aura toujours la même probabilité 1/N; ainsi de suite, à chaque tirage, on aura la probabilité 1/N



- Par contre s'il n'y avait pas de remise:
- Au premier tirage, la probabilité est 1/N
- Au deuxième, la probabilité est 1/(N-1)
-
-
- Au $n^{ème}$, la probabilité est 1/(N-n+1)

– Dans le cas d'un tirage sans remise on ne peut obtenir d'échantillon de taille supérieure à N de la population; on l'appelle donc: **tirage exhaustif**

❑Lorsque la taille de la population est importante par rapport à la taille de l'échantillon, on confond alors le tirage **sans** remise et le tirage **avec** remise

- A priori, la loi $\mathscr{L}$, et en particulier μ et σ^2 sont inconnus.
  - On tire au hasard 9 familles avec remise, et on observe la réalisation de la variable **X** pour les familles tirées. Ce qui revient à observer les réalisations de 9 V.A. *indépendantes et de même loi*.

II- ECHANTILLONNAGE

- Le but de ce paragraphe est d'étudier les liens théoriques existant entre la population et l'échantillon aléatoire prélevé dans cette population.

Définition: Les n V.A. $X_1, X_2, \cdots, X_n$ constituent un échantillon aléatoire simple de la V.A. X si et seulement si $X_1, X_2, \cdots, X_n$ sont indépendantes et de même loi que X.

- Désormais, on appellera **X** la *VA parente.*
- Remarque:

$$E(X_1)=E(X_2)=\cdots=E(X_n)=E(X)=\mu$$
$$V(X_1)=V(X_2)=\cdots=V(X_n)=V(X)=\sigma^2$$

1-L'Echantillonnage aléatoire simple

- -On réalise une étude démographique sur la fécondité chez la femme citadine. Pour ce, on considère la variable aléatoire **X** qui désigne ***le nombre d'enfants*** par famille.
- Soit $\mathscr{L}$ la loi de **X**. L'espérance μ et la variance σ^2 sont deux paramètres de cette loi.
- On note alors **X** $\rightsquigarrow \mathscr{L}(\mu, \sigma^2)$
- Avec μ=E(**X**) et σ^2=V(**X**)

2- La moyenne d'échantillonnage

Il s'agit toujours de l'étude concernant la fécondité chez la femme citadine. On s'intéresse au nombre moyen d'enfants par famille. Pour cela, on prélève 5 échantillons aléatoires et on observe la réalisation des 9 V.A $X_1, X_2, \cdots, X_9$ pour chacun des 5 échantillons:

Echantillon	1	2	3	4	5
X_1	2	1	4	4	0
X_2	1	0	3	4	1
X_3	1	0	3	0	0
X_4	1	4	0	1	2
X_5	3	3	1	0	2
X_6	2	2	2	3	2
X_7	5	5	5	2	4
X_8	2	1	2	4	3
X_9	4	0	2	1	5
$\overline{X}$	2,3	1,8	2,4	2,1	2,1

- **Quelques exemples de statistiques:**

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1 + \cdots + X_n}{n} \text{ moyenne échantillon}$$

$$S_e^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 ; \text{ variance échantillon}$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 ; \text{ quasi - variance}$$

- On remarque que le nombre moyen d'enfants par famille $\overline{X}$ prend des valeurs différentes selon l'échantillon considéré; où :

$$\overline{X} = \frac{1}{9}\sum_{i=1}^{9} X_i = \frac{X_1 + X_2 + \cdots + X_9}{9}$$

- $\overline{X}$ est donc une variable aléatoire; c'est donc une ***statistique.***

I- Propriétés de la statistique $\overline{X}$

- Le tableau précédent nous a permis d'obtenir 5 réalisations de la moyenne d'échantillonnage qu'on note

$$\overline{X}_1 ; \overline{X}_2 ; \cdots ; \overline{X}_5$$

On s'attend à ce que ces 5 valeurs soient proches de la moyenne **μ.**

- **Définition :** Soit $X_1, X_2, \ldots, X_n$ un échantillon aléatoire simple de taille n, et h une fonction de $R^n \rightarrow R$; la variable aléatoire

$$Y = h(X_1, X_2, \ldots, X_n)$$

est appelée une **statistique.**

1- Calcul de $E(\overline{X})$

- **A) Propriété 1:**

Soit $X_1, X_2, \ldots, X_n$ un échantillon aléatoire simple de taille n, relatif à la V.A. parente X. L'espérance de la V.A. $\overline{X}$ est égale à la moyenne de la population **μ** :

$$E(\overline{X}) = E(X) = \mu$$

2- Calcul de $V(\overline{X})$

- **Propriété 2 :**

Soit $X_1, X_2, \ldots, X_n$ un échantillon aléatoire simple de taille n, relatif à la V.A. parente X. La variance de la V.A. $\overline{X}$ est égale à la variance de X divisée par la taille n de l'échantillon:

$$V(\overline{X}) = \frac{V(X)}{n} = \frac{\sigma^2}{n}$$

I- Propriétés de la statistique S_e^2

1- Calcul de $E(S_e^2)$

- **Propriété 3:** Soit $X_1, X_2, \ldots, X_n$ un échantillon aléatoire simple de taille n, relatif à la V.A. parente X de moyenne μ et de variance σ^2. Alors, l'espérance de S_e^2 est égale à:

$$E(S_e^2) = \sigma^2\left(1 - \frac{1}{n}\right)$$

- **Remarque :** $\lim_{n\to\infty} V(\overline{X}) = 0$

- **Exemple**

- Soit $X \rightsquigarrow \mathcal{P}(2)$ la V.A. parente :

$$E(X) = \lambda = 2 = V(X)$$

D'où:

$$E(\overline{X}) = 2 \quad \text{et} \quad V(\overline{X}) = 2/n$$

Remarque:

- L'espérance de la variance d'échantillonnage n'est pas une image parfaite de la variance σ^2:

$$E(S_e^2) \neq \sigma^2$$

- Pour remédier à cet inconvénient, on construit une statistique qui approchera le mieux σ^2

C- La variance d'échantillonnage et la quasi-variance

- On note S_e^2 la variance d'échantillonnage:

$$S_e^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

- On rappelle que S_e^2 est une statistique: c'est une V.A. qui associe une valeur numérique à chaque tirage d'échantillon.

- **Propriété 4 :** Soit $X_1, X_2, \ldots, X_n$ un échantillon aléatoire simple de taille n, relatif à la V.A. parente X de moyenne μ et de variance σ^2. Alors, l'espérance de la quasi-variance est égale à la variance de la population:

$$E(S^2) = \sigma^2$$

Remarque

- On peut montrer également que les variances de S_e^2 et de S^2 tendent toutes les deux vers zéro lorsque la taille de l'échantillon tend vers l'infini :

$$\lim_{n\to\infty} V\left(S_e^2\right) = \lim_{n\to\infty} V\left(S^2\right) = 0$$

- On extrait de cette production un E.A. de taille n. Soit X_1, X_2, …, X_n cet échantillon aléatoire. Alors chaque X_i suit une loi de Bernoulli de paramètre p

$$X_i \rightsquigarrow \mathcal{B}(p)$$

- De plus les X_i sont indépendantes, par conséquent, leur somme S_n suit une loi binomiale :

$$S_n \rightsquigarrow \mathcal{B}(n,p)$$

D- Notion de Fréquence

- Lors d'une production industrielle des pièces mécaniques, on s'intéresse à la **proportion** des pièces défectueuses. Si on note X la V.A. qui prend, avec une probabilité p, la valeur 1 si la pièce est défectueuse et qui prend 0, avec la probabilité 1-p sinon. on note alors :

- S_n représente le nombre de pièces défectueuses dans l'échantillon :

$$P(S_n = k) = C_n^k p^k q^{n-k} \; ; k = 0,1,\ldots,n$$

- Définissons la fréquence F comme étant la proportion de pièces défectueuses dans l'échantillon :

$$F = \frac{S_n}{n} = \frac{X_1 + X_2 + \cdots X_n}{n}$$

$$X = \begin{cases} 1 \text{ avec la proba. } p \text{ si pièce défectueuse} \\ 0 \text{ avec proba. } 1-p \text{ si pièce conforme} \end{cases}$$

- D'où :

$$X \rightsquigarrow \mathcal{B}(p)$$

- Si x est une valeur numérique que peut prendre la statistique F, alors :

$$x \in \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\right\}$$

- Et la distribution de F est donnée par :

$$P(F = x) = P(S_n = nx) = C_n^{nx} p^{nx} q^{n-nx}$$

•**Remarque :**Les propriétés de F sont déduites de celles de $\overline{X}$ (puisque F est un cas particulier de $\overline{X}$)

•**Propriété 6 :** Soit la V.A. parente X qui suit une loi de Bernoulli de paramètre p. Les moments de la statistique ***F*** sont comme suit :

$$E(F)=p \text{ et } V(F)=\frac{pq}{n}$$

- **Remarque :**
  - On peut redéfinir la convergence en probabilité comme suit :

$$\forall \varepsilon>0;\ \lim_{n\to\infty} P\left\{\left|X_n-X\right|>\varepsilon\right\}=0$$

  - La V.A. X peut être une *constante*; par exemple, lorsqu'on étudie la convergence d'une suite de statistiques vers la valeur vraie d'un paramètre

E- Théorèmes fondamentaux de la Statistique

I- Convergence en probabilité

- **Définition :** La suite aléatoire X_1, X_2, …, X_n,… converge en probabilité vers la V.A. X, appelée limite de la suite, si :

$$\forall \varepsilon>0;\ \lim_{n\to\infty} P\left\{\left|X_n-X\right|<\varepsilon\right\}=1$$

- Ce que l'on note:

$$X_n \xrightarrow{P} X$$

II- Convergence en LOI

- **Définition :** La suite aléatoire X_1, X_2, …, X_n,… converge en loi vers la V.A. X, appelée limite de la suite si

$$\lim_{n\to\infty} F_n(x)=F(x);$$

en tout point de continuité ***x*** de ***F(x)***

- Et on note :

$$X_n \xrightarrow{L} X$$

- Où ***F(x)*** est la fonction de répartition de X

Convergence de la loi de Poisson vers la loi Normale

- **Propriété 10 :** Soit une V.A. Z normale centrée réduite: $Z \rightsquigarrow \mathcal{N}(0,1)$ et soit une suite de V.A. X_1, X_2, …, X_n,… telles que: $X_n \rightsquigarrow \mathcal{P}(\lambda_n)$ avec $\lim_{n\to\infty}\lambda_n=+\infty$

Alors;

$$\frac{X_n-\lambda_n}{\sqrt{\lambda_n}} \xrightarrow{L} Z$$

Remarque:

- Si λ est suffisamment grand $(\lambda \geq 15)$,

Alors :

$$\mathcal{P}(\lambda) \approx \mathcal{N}(\lambda, \lambda)$$

Convergence de la loi de Khi-deux vers la loi Normale

Rappel sur la loi de Khi-deux: χ^2

Convergence de la loi Binomiale vers la loi Normale

- **Propriété 11 :** Soit une V.A. Z normale centrée réduite: $Z \rightsquigarrow \mathcal{N}(0,1)$ et soit une suite de V.A. $X_1, X_2, \ldots, X_n, \ldots$ telles que: $X_n \rightsquigarrow \mathcal{B}(n, p)$

Alors;

$$\frac{X_n - np}{\sqrt{npq}} \xrightarrow{L} Z$$

- **Définition :** Soit une suite de n V.A. $Z_1, Z_2, \ldots, Z_n$ indépendantes et de même loi normale centrée réduite: $Z_i \rightsquigarrow \mathcal{N}(0,1)$

On appelle loi de Khi-deux à n degrés de liberté, la loi suivie par la somme des carrés des V.A. Z_i; on note :

$$\sum_{i=1}^{n} Z_i^2 \rightsquigarrow \chi^2(n)$$

Remarque:

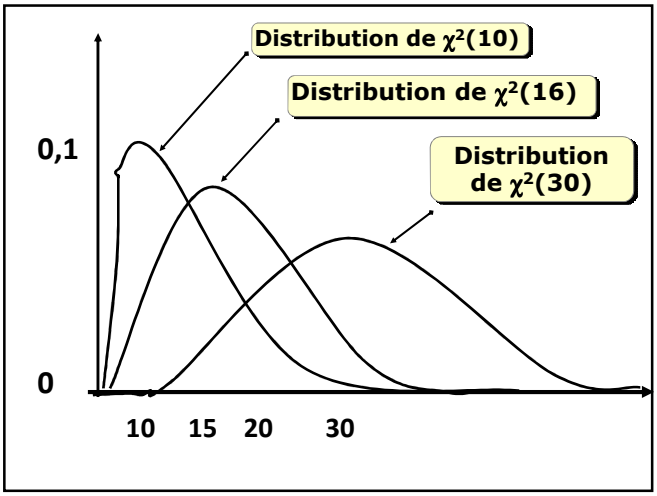
- Si $n \geq 30$ et $np \geq 5$ et $nq \geq 5$

Alors :

$$\mathcal{B}(n, p) \approx \mathcal{N}(np, npq)$$

b)- Distribution d'une Khi-deux

- La distribution de χ^2 est dissymétrique. Elle est continue et elle dépend du nombre de degrés de liberté n. Pour des valeurs différentes de n, on obtient des distributions différentes. Lorsque n augmente, la loi de χ^2 tend lentement vers la loi normale



- **Propriété 12:** Approximation de fisher
- Soit $X \rightsquigarrow \chi^2(n)$, si $n > 30$; alors :

$$\sqrt{2X} - \sqrt{2n-1} \approx \mathcal{N}(0,1)$$

C)- Les moments de Khi-deux

- Soit $X \rightsquigarrow \chi^2(n)$; alors :

$$E(X) = n \quad \text{et} \quad V(X) = 2n$$

Convergence de la loi de Student vers la loi Normale

Rappel sur les lois de Student : t et de Fisher : F

d)- Une propriété de Khi-deux

- La somme de deux Khi-deux indépendantes de d.d.l. respectifs n et m est une Khi-deux de d.d.l. n+m ; on note :

$$X \rightsquigarrow \chi^2(n) \quad \text{et} \quad Y \rightsquigarrow \chi^2(m) \quad \text{où } X \perp\!\!\!\perp Y \Rightarrow X+Y \rightsquigarrow \chi^2(n+m)$$

- **Définition 1**: Soient X une V.A. de khi-deux à n d.d.l. et Y une V.A. de khi-deux à m d.d.l. avec X et Y indépendantes; on définit alors la V.A. de Fisher par le rapport des rapports des deux V.A. de khi-deux par leurs d.d.l. respectifs n et m et on note :

$$X \rightsquigarrow \chi^2(n)$$

- $$\text{et} \quad \text{où } X \perp\!\!\!\perp Y \Rightarrow \frac{X/n}{Y/m} \rightsquigarrow \mathcal{F}(n,m)$$

$$Y \rightsquigarrow \chi^2(m)$$

- $\mathcal{F}(n, m)$ est la notation de la loi de Fisher à (n, m) d.d.l. Sa densité étant dissymétrique, étalée à droite.

- **Propriété 14:** moments de la loi de Fisher

Soit $F \rightsquigarrow \mathcal{F}(n, m)$,alors

$$E(F) = \frac{m}{m-2} \quad si \quad m > 2$$

$$V(F) = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)} \quad si \quad m > 4$$

Distribution de la loi de Student

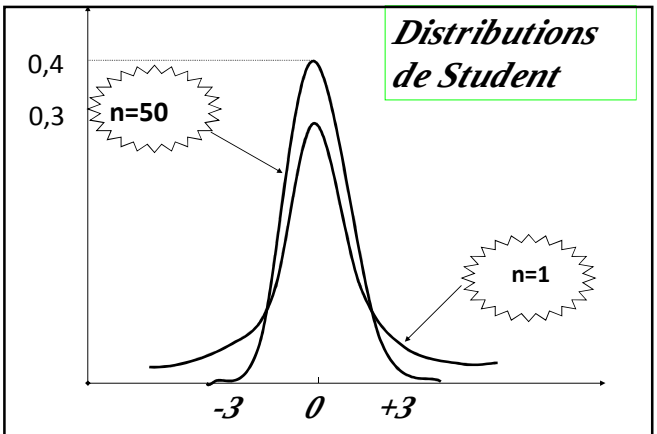
- Les distributions de Student sont continues et symétriques et dépendent d'un paramètre n. A des valeurs différentes de n, correspond différentes distributions de Student. Lorsque n tend vers l'infini, la loi de student tend vers la loi normale.

- **Remarque :**

➢ Si $F_{n,m,p}$ est le fractile d'ordre p de la V.A. de Fisher à (n, m) d.d.l, alors :

$$F_{n,m,p} = \frac{1}{F_{m,n,1-p}}$$

➢ Cette loi joue un grand rôle en statistique (loi du rapport des variances de deux échantillons indépendants)



- **Définition :** Soit Z une V.A. suivant la loi normale centrée réduite: $Z \rightsquigarrow \mathcal{N}(0,1)$ et soit X une V.A. de Khi-deux à ν degrés de liberté: $X \rightsquigarrow \chi^2(\nu)$ et indépendante de Z. On définit alors la V.A. T suivant la loi de student à ν degrés de liberté, notée t_ν:

$$T = \frac{Z}{\sqrt{X/\nu}} \rightsquigarrow t_\nu$$

Propriété 15: moments de la loi de Student

Soit $T \rightsquigarrow t(n)$,alors

$$E(T) = 0 \quad si \quad n > 1$$

$$V(T) = \frac{n}{(n-2)} \quad si \quad n > 2$$

$$\mu_3 = 0 \quad si \quad n > 3$$

Propriétés de Student

- ✓ $T^2=F(1,n)$
- ✓ Soit $T \rightsquigarrow t(n)$, alors si n >120

$$T \approx \mathcal{N}(0,1)$$

Loi des Grands Nombres (loi faible)

- **Théorème 1:** Loi des grands nombres
- Soit $X_1, X_2, \ldots, X_n$ un échantillon aléatoire simple de taille n, relatif à la V.A. parente X de moyenne μ ; alors :

$$\overline{X} \xrightarrow{P} \mu$$

- **Remarque**: Ce résultat reste vrai quelque soit la loi de X. donc en particulier pour la loi de Bernoulli

- **Corollaire 1:** Soit la V.A. parente X suivant une loi de Bernoulli de paramètre p ; alors :

$$F \xrightarrow{P} p$$

- **Remarque**: Cela veut dire que la fréquence d'un événement converge vers sa probabilité
- La démonstration du Théorème1 découle du lemme suivant

Lemme : Inégalité de Bienaymé-Tchebicheff

Si **ε** désigne un réel strictement positif, et X une V.A. d'écart type **σ**, alors

$$P\{|X-E(X)| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$$

- ***Indication pour démonstration***:
- voir dans le cas d'une v.a. discrète et poser $Y=(X-EX)^2$ et calculer EY

- **Remarque:**
  - La loi des grands nombres peut être généralisée:
- **Théorème 2:** Soit $T=T(X_1, X_2, \ldots, X_n)$ une statistique telle que :

$$\lim_{n\to\infty} E(T)=\theta \quad et \quad \lim_{n\to\infty} V(T)=0$$

alors : $T \xrightarrow{P} \theta$

- **Exemple :**
- 1)-

$$\lim_{n\to\infty} E\left(S_e^2\right)=\sigma^2 \quad et \quad \lim_{n\to\infty} V\left(S_e^2\right)=0$$

Alors $S_e^2 \xrightarrow{P} \sigma^2$

- 2)-

$$\lim_{n\to\infty} E\left(S^2\right)=\sigma^2 \quad et \quad \lim_{n\to\infty} V\left(S^2\right)=0$$

Alors :

$$S^2 \xrightarrow{P} \sigma^2$$

Cas des échantillons Aléatoires issus d'une population Normale

- Lorsque l'E.A. est relatif à une loi normale, on obtient des propriétés plus intéressantes pour les statistiques $\overline{X}$ et S^2
- **Propriété 16:** Soit $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$, alors

$$\overline{X} \rightsquigarrow \mathcal{N}(\mu\ ,\ \sigma^2/n)$$

Théorème CENTRAL LIMITE (T.C.L.)

- **Théorème 3:** (T.C.L. 1ère *formulation*)
- Soit $X_1, X_2, \ldots, X_n$ un échantillon aléatoire simple de taille n, relatif à la V.A. parente X de moyenne μ et de variance σ^2. Alors,
- $$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n}\,\frac{\overline{X}-\mu}{\sigma} \xrightarrow{L} \mathcal{N}(0,1)$$

- **Remarque:** Le T.C.L. s'applique à toute V.A. quelque soit sa loi; il fournit une propriété asymptotique.
- Alors que la **propriété 16** ne s'applique qu'aux V.A. suivants une loi Normale et quelque soit la taille n de l'échantillon.

- **Théorème 4:** (T.C.L. 2ème *formulation*)
- Pour une taille n assez grande (en pratique $n \geq 30$), on a:
- $$\overline{X} \approx \mathcal{N}\left(\mu, \sigma^2/n\right)$$

- **Corollaire 2:** Soit la V.A. parente X suivant une loi de Bernoulli de paramètre p ; Si $n \geq 30$ et $np \geq 5$ et $nq \geq 5$ alors on a :

$$F \approx \mathcal{N}\left(p, \frac{pq}{n}\right)$$

Propriété 17: Sous l'hypothèse de normalité,

$$\frac{(n-1)}{\sigma^2} S^2 \rightsquigarrow \chi^2(n-1)$$

Propriété 18: Sous l'hypothèse de normalité:

$$\sqrt{n}\,\frac{\overline{X}-\mu}{S} \rightsquigarrow \mathbf{t}(n-1)$$

Remarque: Sous l'hypothèse de normalité, et si la moyenne de la population est connue, on a:

$$\frac{n}{\sigma^2}\tilde{S}^2 \rightsquigarrow \chi^2(n)$$

$$\text{Où } \tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

- Lorsque le tirage est sans remise, toutes les propriétés qu'on a énoncé (ou presque), concernant les statistiques , $\overline{X}, F, S_e^2$ et S^2 ne sont plus valables. En effet, on peut montrer qu'on a :

$$\begin{cases} E(\overline{X}) = \mu \\ V(\overline{X}) = \frac{\sigma^2}{n}\frac{N-n}{N-1} \end{cases}$$

$$\begin{cases} E(F) = p \\ V(F) = \frac{pq}{n}\frac{N-n}{N-1} \end{cases}$$

Remarque: Cas des petits échantillons

- Lorsque n est petit, on ne peut utiliser le T.C.L. et par conséquent on ne peut obtenir une loi pour $\overline{X}$. Or si on ajoute l'hypothèse de normalité, on obtient des résultats importants quelque soit ***n***.

$$\begin{cases} E(S_e^2) = \sigma^2\left(\frac{n-1}{n}\right)\left(\frac{N}{N-1}\right) \\ E(S^2) = \sigma^2\frac{N}{N-1} \end{cases}$$

- Remarque: Si n est négligeable devant N (la taille de la population finie), le tirage sans remise devient équivalent à un tirage avec remise. Dans la pratique, ceci prend effet lorsque: $N \geq 20n$

F- Cas des échantillons exhaustifs (T.S.R.)

L' Estimation

Chapitre II

INTRODUCTION

- Après avoir prélevé l'échantillon, et étudié les distributions d'échantillonnage, on peut alors généraliser, à la population, les résultats expérimentaux obtenus à partir de l'échantillon; c'est ce qu'on appelle l'inférence statistique.

- ✓ L'estimation consiste en l'évaluation d'un paramètre de la population à partir de l'observation d'un E.A.
- ✓ La théorie de l'estimation se divise en deux parties:

1) ***L'estimation ponctuelle***: permet d'obtenir une valeur unique calculée à partir d'un E.A., valeur qui sera prise comme estimation du paramètre inconnu.

2) ***L'estimation par intervalle***: permet de déterminer un intervalle qui, avec une grande probabilité fixée a priori, contient la valeur vraie du paramètre inconnu.

- ***Exemple:*** On dit que 53% de la population favorise le candidat A avec une marge d'erreur de 1% et avec un niveau de confiance de 95%. Ce qui signifie que la proportion d'électeurs favorisant A se situe, avec une probabilité de 95%, entre 52% et 54%.

A- L' ESTIMATION PONCTUELLE

- I- **Définitions**
  - Définition1: Soit $X_1, X_2, \ldots, X_n$ un E.A.S relatif à la V.A. parente X de loi $\mathscr{L}(\theta)$.

    On appelle estimateur du paramètre θ toute statistique utilisée dans le but d'approcher la valeur inconnue de θ. Un estimateur est donc une Variable Aléatoire.

  - **Définition2**: Une estimation du paramètre θ est une réalisation d'un estimateur de ce paramètre. Une estimation est donc une valeur numérique.
  - **Exemples:**

Si $X \rightsquigarrow \mathscr{L}(\mu) \longrightarrow$ estimateur $\overline{X} \rightarrow$ estimation $\overline{x}$

Si $X \rightsquigarrow \mathscr{L}(p) \longrightarrow$ estimateur $F \rightarrow$ estimation f

Si $X \rightsquigarrow \mathscr{L}(\sigma^2) \longrightarrow$ estimateur $\begin{cases} S_e^2 \\ S^2 \end{cases} \rightarrow$ estimations $\begin{cases} s_e^2 \\ s^2 \end{cases}$

- II- **Propriétés des estimateurs**

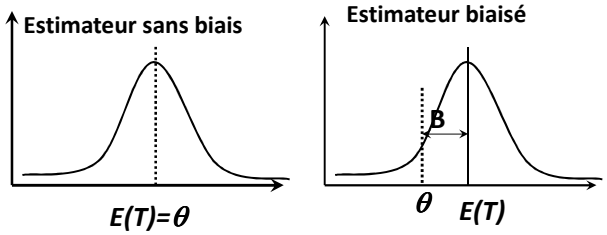
- **1-Estimateurs sans biais**

- **Définition:** On appelle estimateur sans biais du paramètre θ toute statistique $T=T(X_1, X_2, \ldots, X_n)$ telle que:

$$E(T)=\theta$$

- **Remarque 1:** Si T est biaisé, le biais serai alors:

$$B = E(T) - \theta$$



2-Estimateurs Convergents

- **Définition:** On appelle estimateur Convergent du paramètre θ toute statistique $T=T(X_1, X_2, \ldots, X_n)$ telle que:

$$T \xrightarrow{P} \theta$$

- ***Exemples***:

$$\begin{cases} \overline{X} \text{ est un estimateur sans biais de } \mu \\ S^2 \text{ est un estimateur sans biais de } \sigma^2 \\ S_e^2 \text{ est un estimateur biaisé de } \sigma^2 \end{cases}$$

$$B\left(S_e^2\right) = E\left(S_e^2\right) - \sigma^2 = \left(\frac{n-1}{n}\right)\sigma^2 - \sigma^2 = \frac{-\sigma^2}{n}$$

- ***Remarque*** : La loi des grands nombres implique que $\overline{X}$ est un estimateur convergent de μ. De même que F est un E.C. de p

- ***Définition:*** On appelle estimateur asymptotiquement sans biais, du paramètre θ, toute statistique $T=T(X_1, X_2, \ldots, X_n)$ telle que:

$$\lim_{n\to\infty} E(T) = \theta$$

Exemple:

S_e^2 est un estimateur asymptotiquement sans biais de σ^2 :

$$\lim_{n\to\infty} E\left(S_e^2\right) = \lim_{n\to\infty}\left(1 - \frac{1}{n}\right)\sigma^2 = \sigma^2$$

Théorème : **L.G.N. généralisée**

- Si T est un estimateur sans biais ou asymptotiquement sans biais de θ, et si sa variance tend vers zéro lorsque n tend vers l'infini, alors T est un estimateur convergent de θ :
- Si $\lim_{n\to\infty} E(T) = \theta \quad et \quad \lim_{n\to\infty} V(T) = 0$
- Alors $T \xrightarrow{P} \theta$

Exemple :

$$\lim_{n\to\infty} E\left(S_e^2\right)=\sigma^2 \quad et \quad \lim_{n\to\infty} V\left(S_e^2\right)=0$$

Alors

S_e^2 est un estimateur convrgent de σ^2

- De même on montre que S^2 est un E.C. de σ^2

- ***Question***: $\underset{T\in\{E.S.B.\}}{Min}\ V(T)$?

- ***Définition***: On appelle Quantité d'information de Fisher de θ, la quantité $I_n(\theta)$:

$$I_n(\theta)=E\left[\left(\frac{\partial}{\partial\theta}Log\ f(X_1,X_2,\ldots,X_n,\theta)\right)^2\right]$$

- Où $I_n(\theta)$ dépend de n et de la loi de X mais ne dépend pas de l'estimateur T

3-Estimateurs Efficaces:

Entre deux estimateurs sans biais, on préfère utiliser celui qui a la variance la plus petite.

- **Définition:** Soient T_1 *et* T_2 deux estimateurs sans biais du même paramètre θ et basés sur le même échantillon. On dit que T_1 *est plus précis que* T_2 si:

$$V(T_1)\le V(T_2)$$

- ***Remarques*** : Soit $f(x,\theta)$ la densité de la loi de X, alors $I_n(\theta)$ s'écrit aussi (sous certaines conditions qu'on supposera toujours vérifiées):

1-
$$I_n(\theta)=nE\left[\left(\frac{\partial}{\partial\theta}Log\ f(X,\theta)\right)^2\right]$$

2-
$$I_n(\theta)=nI_1(\theta)$$

3-
$$I_n(\theta)=-nE\left(\frac{\partial^2}{\partial\theta^2}Log\ f(X,\theta)\right)$$

Exemple : Montrons que $\overline{X}$ est plus précis que $\Gamma=\frac{X_1+X_2}{2}$. On a d'abord:

$E\left(\overline{X}\right)=E(\Gamma)=\mu$ donc $\overline{X}$ *et* Γ sont deux E.S.B. de μ.

$$V(\Gamma)=\frac{1}{4}(V(X_1)+V(X_2))=\frac{1}{2}V(X)=\frac{\sigma^2}{2}$$

Or $V\left(\overline{X}\right)=\frac{\sigma^2}{n}$

D'où

$$V\left(\overline{X}\right)\le V(\Gamma)\ pour\ n\ge 2$$

4- Le rapport $B_{CR}(\theta)=\frac{1}{I_n(\theta)}$ est appelé

Borne de *CRAMER-RAO*

- ***Propriété*** : **Inégalité de *CRAMER-RAO***

Si *T* est un Estimateur sans biais **(E.S.B.)** du paramètre θ, alors :

$$V(T) \geq B_{CR}(\theta)$$

III Recherche d'estimateurs

- Soit (x_1, x_2, …, x_n) la réalisation d'un échantillon aléatoire X_1, X_2, …, X_n, relatif à la V.A. parente X de loi P_θ. La probabilité d'obtenir une telle réalisation est :

$$P_\theta(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \prod_{i=1}^{n} P_\theta(X_i = x_i)$$

- ***Définition*** : un E.S.B. T de θ, est dit efficace si sa variance est égale à la borne de Cramer-Rao:

$$V(T) = B_{CR}(\theta)$$

- ***Définition:*** un E.S.B. T de θ, est dit asymptotiquement efficace si:

$$\lim_{n \to \infty} \frac{V(T)}{B_{CR}(\theta)} = 1$$

- ***Définition*** : On appelle Estimateur de Maximum de Vraisemblance (E.M.V.) pour θ, la valeur $\hat{\theta}$ de θ, qui maximise cette probabilité.

- En pratique, il revient au même de chercher $\hat{\theta}$ qui maximise le logarithme de la probabilité, soit :

$$L(\theta) = \sum_{i=1}^{n} Log\left[P_\theta(X_i = x_i)\right]$$

Exemple : Si X $\rightsquigarrow \mathcal{N}(\mu, \sigma^2)$, alors

$$I_n(\mu) = \frac{n}{\sigma^2} \text{ et } V(\overline{X}) = \frac{\sigma^2}{n}$$

Or $\overline{X}$ E.S.B. de μ, d'où il est efficace de μ.

$$I_n(\sigma^2) = \frac{n}{2\sigma^4} \text{ et } V(S^2) = \frac{2\sigma^4}{n-1}$$

S^2 est asymptotiquement efficace de σ^2 car $(n/n-1) \longrightarrow 1$

- ***Remarque*** :
- $\hat{\theta}$ maximise $L(\theta)$ ssi $\begin{cases} \frac{\partial L}{\partial \theta}(\hat{\theta}) = 0 \\ \frac{\partial^2 L}{\partial \theta^2}(\hat{\theta}) < 0 \end{cases}$
- ***Exemple*** :
- Si X $\rightsquigarrow \mathcal{B}$ (p) $\Longrightarrow P_p(X_i = x_i) = p^{x_i} q^{1-x_i}$
- D'où

$$L(p) = \sum_{i=1}^{n} \left[x_i Logp + (1 - x_i) Log(1 - p)\right]$$

$$\frac{\partial L}{\partial p}=\sum_{i=1}^{n}\left(\frac{x_i}{p}-(1-x_i)\frac{1}{1-p}\right)=\frac{1}{p}\sum_{i=1}^{n}x_i-\frac{n-\sum_{i=1}^{n}x_i}{1-p}$$

$$\frac{\partial L}{\partial p}=0\Leftrightarrow p\sum_{i=1}^{n}x_i=n\Leftrightarrow \hat{p}=\frac{\sum_{i=1}^{n}X_i}{n}=F$$

Vérifiez ensuite que $\frac{\partial^2 L}{\partial\theta^2}(\hat{p})<0$

On conclu que F est bien un E.M.V. pour p

B- Estimation Par Intervalle de Confiance

–**Définition**: On dit que (C_1,C_2) est un intervalle de confiance au niveau 1-α pour le paramètre θ si on a :

$$P\left((C_1;C_2)\ni\theta\right)=P\left(C_1\le\theta\le C_2\right)=1-\alpha$$

–Les bornes C_1 et C_2 de l'intervalle sont des statistiques relatives à l' E.A.

IV Cas Exhaustif

1-Cas de la moyenne : $\theta=\mu$

Lorsque le tirage est sans remise, $\overline{X}$ garde la même espérance mais pas la même variance:

$$E\left(\overline{X}\right)=\mu \quad \text{et} \quad V\left(\overline{X}\right)=\frac{\sigma^2}{n}\frac{N-n}{N-1}$$

Remarque : $V\left(\overline{X}_{TSR}\right)\le V\left(\overline{X}_{TAR}\right)$

- 1-α est appelé niveau de confiance de l'intervalle (C_1,C_2).
- Plus α est petit et plus l'intervalle de confiance est grand. Généralement, on considère des intervalles à risques symétriques:

$$P\left(\theta>C_2\right)=P\left(\theta<C_1\right)=\frac{\alpha}{2}$$

2-Cas de la variance : $\theta=\sigma^2$

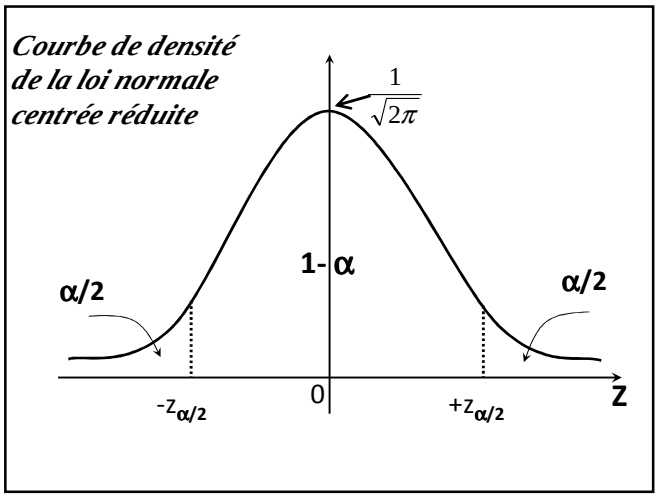
- Si le **TAR** S^2 est un E.S.B. de σ^2
- Mais lorsque le **TSR**, on a:

$$E\left(\frac{N-1}{N}S^2\right)=\sigma^2$$

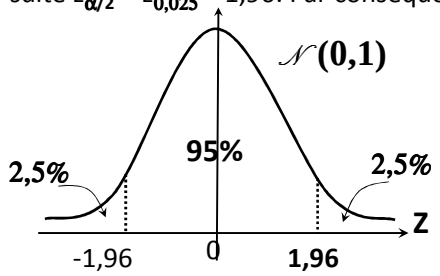
$\frac{N-1}{N}S^2$ est donc un ESB de σ^2

I- Intervalle de Confiance pour une proportion: $I_c(p)$

- **1- Construction de l' $I_c(p)$:** Afin d'estimer une proportion p de « Succès » par exemple; on utilise la fréquence F qui est un estimateur sans biais convergent et efficace du paramètre p. De plus, on a (d'après le T.C.L.)
- $$Z=\frac{F-p}{\sqrt{\frac{pq}{n}}}\approx\mathcal{N}(0,1)$$
- dés que $n\ge 30$ et $np\ge 5$ et $nq\ge 5$



- **Exemple**: pour un niveau de confiance 95%; α=5% et par suite $z_{\alpha/2} = z_{0,025} = 1,96$. Par conséquent:



$$I_C(p) = \left[F - 1,96\sqrt{\frac{pq}{n}} \; ; \; F + 1,96\sqrt{\frac{pq}{n}} \right]$$

- D'où

$$P\left(-z_{\alpha/2} < Z < +z_{\alpha/2}\right) = 1-\alpha$$

- Remplaçons Z par son expression:

$$P\left(-z_{\alpha/2} < \frac{F-p}{\sqrt{\frac{pq}{n}}} < +z_{\alpha/2} \right) = 1-\alpha$$

- Une réalisation de cet intervalle est :

$$I_C(p) = \left[f - 1,96\sqrt{\frac{pq}{n}} \; ; \; f + 1,96\sqrt{\frac{pq}{n}} \right]$$

- Problème: $\sqrt{pq/n}$ est inconnue car p est inconnu. On utilise alors l'approximation suivante, pour n assez grande :

$$\sqrt{pq/n} \cong \sqrt{f(1-f)/n}$$

$$P\left(F - z_{\alpha/2}\sqrt{\frac{pq}{n}} < p < F + z_{\alpha/2}\sqrt{\frac{pq}{n}} \right) = 1-\alpha$$

•Donc, on obtient comme intervalle de confiance pour la proportion au niveau 1-α

$$I_C(p) = \left[\underbrace{F - z_{\alpha/2}\sqrt{\frac{pq}{n}}}_{C_1} \; ; \; \underbrace{F + z_{\alpha/2}\sqrt{\frac{pq}{n}}}_{C_2} \right]$$

- **Propriété:** Soit $X_1, X_2, \ldots, X_n$ un échantillon aléatoire simple de taille n, relatif à la V.A. parente X suivant une loi de Bernoulli de paramètre p inconnu. Un intervalle de confiance au niveau 1-α pour p est comme suit :

$$I_C(p) = \left[f - z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}} \; ; \; f + z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}} \right]$$

- Dés que :

$$n \geq 30 \text{ et } nf \geq 5 \text{ et } n(1-f) \geq 5$$

- **2- Précision d'une estimation par intervalle de confiance:** Pour un α donné, on veut déterminer la taille nécessaire de l'échantillon pour atteindre une certaine précision de l'estimation.
  - Notons *a(α,n)* l'amplitude de $I_C(p)$:

$$a(\alpha,n)=\left(f+z_{\alpha/2}\sqrt{\frac{pq}{n}}\right)-\left(f-z_{\alpha/2}\sqrt{\frac{pq}{n}}\right)$$
$$=2z_{\alpha/2}\sqrt{pq/n}$$

- **Exercice:** Parmi un E.A. de 250 électeurs, 108 déclarent vouloir voter pour le président sortant. Tandis que les autres voteront pour l'autre candidat. Donner un $I_C(p)$ au niveaux 90 et 95%
- Combien faudrait-il interroger d'électeurs pour que l'erreur d'estimation ne dépasse pas 2% ?

• Réponse

- 1)-

$$X=\begin{cases}1 \text{ avec proba. } p & \text{si vote pour A}\\ 0 \text{ avec proba. } 1-p & \text{si vote pour B}\end{cases}$$

- D'où : $X \rightsquigarrow \mathscr{B}(p)$

- On appelle erreur d'estimation la demi-longueur de I_c

$$\frac{a(\alpha,n)}{2}=z_{\alpha/2}\sqrt{pq/n}$$

- Or p est inconnu, on utilise donc la majoration:

$$\sqrt{p(1-p)}\le 1/2$$

- On obtient donc

$$z_{\alpha/2}\sqrt{pq/n}\le z_{\alpha/2}\frac{1}{2\sqrt{n}}$$

- ***p*** représente la proportion des votants en faveur de A dans toute la population des électeurs.
- Une estimation ponctuelle de p est donnée, grâce à la réalisation de l'E.A., par la fréquence:

$$f=\frac{n_1}{n}=\frac{108}{250}=0{,}432$$

- Cond$^{\theta}$ TCL:

$$n=250\ge 30 \text{ et } nf=108\ge 5 \text{ et } 142\ge 5$$

Or α=10%, d'où $z_{\alpha/2}=z_{0,05}=1{,}645$(loi normale C.R)

- Pour que l'erreur d'estimation soit inférieure ou égale à e:

$$\frac{z_{\alpha/2}}{2\sqrt{n}}\le e$$

- Il faut que n soit:

$$n\ge\frac{1}{4}\left(\frac{z_{\alpha/2}}{e}\right)^2$$

- **Remarque:** Plus l'erreur d'estimation est petite et plus la précision est grande.

$$I_C(p)=\left[f-z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}\,;\,f+z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}\right]$$
$$=[0{,}432-1{,}645\times 0{,}031;\,0{,}432+1{,}645\times 0{,}031]$$
$$=[0{,}381;\,0{,}483]$$

- 2)- α=5% ⇒ $z_{0,025}=1{,}96$

$$I_C(p)=[0{,}371\ ;\ 0{,}493]$$

- 3)-

$$n \geq \frac{1}{4}\left(\frac{z_{\alpha/2}}{e}\right)^2 = \frac{1}{4}\left(\frac{1{,}96}{0{,}02}\right)^2 = 2401$$

05/01/2004

- D'où :

$$I_C(\mu) = \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} ; \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

- Cet intervalle n'a de sens que si σ^2 est connue.
- Si σ^2 est inconnue, on l'estime alors par S^2 et on obtient la V.A.(grâce à l'hypothèse de Normalité) :
- $Z = \sqrt{n}\,\frac{\bar{X} - \mu}{S} \rightsquigarrow$ **t(n-1)**

II- Intervalle de Confiance pour la moyenne: $I_C(\mu)$

- **1)- $I_C(\mu)$ dans le cas d'un population Normalement distribuée:**
- Dans ce cas la V.A. parente X suit :
- $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$.
- Or, on sait que $\bar{X}$ est un bon estimateur de la moyenne μ :
- $\bar{X} \rightsquigarrow \mathcal{N}(\mu, \sigma^2/n)$.

- Notre intervalle de confiance devient :

$$I_C(\mu) = \left[\bar{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} ; \bar{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\right]$$

- Où $t_{n-1,\alpha/2}$ est la valeur critique d'ordre $\alpha/2$ lue dans la table de student à n-1 d.d.l.

- Posons
- $$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow \mathcal{N}(0, 1).$$
- D'où

$$1 - \alpha = P\left(-z_{\alpha/2} < Z < +z_{\alpha/2}\right)$$

$$= P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

- **Propriété:** Soit la V.A $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$.
- **1)-** Si σ^2 est **connue**, alors un intervalle de confiance au niveau 1-α pour μ est :

$$I_C(\mu) = \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} ; \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

- **2)-** Si σ^2 est **inconnue**, alors un intervalle de confiance au niveau 1-α pour μ est :

$$I_C(\mu) = \left[\bar{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} ; \bar{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\right]$$

- **2)- $I_c(\mu)$ dans le cas d'un population quelconque:**

- La variance de X peut-être considérée comme une mesure de l'incertitude de l'actif et donc comme une mesure de son risque.
- Soit P la population des actions des entreprises du secteur micro-informatique. X est le taux de rendement de ses actions; on suppose :
- $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$.
- On prélève au hasard 30 actions et construisons un intervalle de confiance pour σ^2

- **Propriété:** Soit $X_1, X_2, \ldots, X_n$ un échantillon aléatoire relatif à $X \rightsquigarrow LQ(\mu, \sigma^2)$.

1)- Si σ^2 est connue, et si $n \geq 30$ alors un intervalle de confiance au niveau 1-α de μ

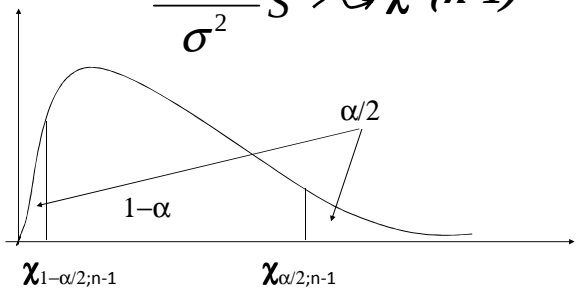
$$I_C(\mu) = \left[\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} ; \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

2)- Si σ^2 est inconnue, et si $n \geq 50$ alors notre intervalle pour μ devient :

$$I_C(\mu) = \left[\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}} ; \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}} \right]$$

Si μ est inconnue, on utilise alors S^2 et on obtient la V.A.(grâce à l'hypothèse de Normalité)

$$\frac{(n-1)}{\sigma^2} S^2 \rightsquigarrow \chi^2(n-1)$$



12/01/2004

III- Intervalle de Confiance pour la variance: $I_c(\sigma^2)$

- Pour montrer le grand intérêt que présente la variance, on prend l'exemple où X est le taux de rendement d'un actif financier. Celui-ci est définit comme suit:

$$R_t = \frac{P_t + dt - P_{t-1}}{P_{t-1}}$$

- Où P_t est le prix de la période t, et dt le dividende de la même période

- D'où :

$$I_C(\sigma^2) = \left[\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\chi^2_{n-1;\alpha/2}} ; \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\chi^2_{n-1;1-\alpha/2}} \right]$$

- ou

$$I_C(\sigma^2) = \left[\frac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}} ; \frac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}} \right]$$

IV- Intervalle de Confiance pour la différence des moyenne

•1)- $I_c(\mu_1 - \mu_2)$ dans le cas de 2 populations Normalement distribuées:

•Soient P_1 et P_2 2 populations que l'on étudie selon la V.A. parente X de loi sur P_1 $\mathcal{N}(\mu_1, \sigma_1^2)$ et $\mathcal{N}(\mu_2, \sigma_2^2)$ sur P_2.

•On veut estimer la différence des moyennes $\mu_1 - \mu_2$.

- Si les deux variances σ_1^2 et σ_2^2 sont connues :

$$I_C(\mu_1-\mu_2)=\left[\bar{x}_1-\bar{x}_2-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}};\ \bar{x}_1-\bar{x}_2+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right]$$

- Si les variances sont inconnues mais égales

$$\sigma_1^2=\sigma_2^2=\sigma^2$$

- Pour cela, on prélève 2 échantillons indépendamment de tailles n_1 et n_2 respectivement de P_1 et P_2
- Et on définie alors 2 estimateurs $\overline{X}_1$et $\overline{X}_2$ respectivement de μ_1 et μ_2

On propose alors $\overline{X}_1 - \overline{X}_2$ estimateur de $\mu_1 - \mu_2$:

$$\mu_{\overline{X}_1-\overline{X}_2}=E\left(\overline{X}_1-\overline{X}_2\right)=E\left(\overline{X}_1\right)-E\left(\overline{X}_2\right)$$
$$=\mu_1-\mu_2$$

On les estime alors par $\hat{S}^2$, où

$$\hat{S}^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{(n_1-1)+(n_2-1)}$$

- Et donc on construit une V.A. de Student:

$$T=\frac{\left(\overline{X}_1-\overline{X}_2\right)-(\mu_1-\mu_2)}{\hat{S}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \rightsquigarrow t(n_1+n_2-2)$$

$$\sigma^2_{\overline{X}_1-\overline{X}_2}=\sigma^2_{\overline{X}_1}+\sigma^2_{\overline{X}_2}=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$$

- Or $\overline{X}_1 \rightsquigarrow \mathcal{N}(\mu_1, \sigma_1^2/n_1)$.
- Et $\overline{X}_2 \rightsquigarrow \mathcal{N}(\mu_2, \sigma_2^2/n_2)$.
- D'où

$$\overline{X}_1-\overline{X}_2 \rightsquigarrow \mathcal{N}(\mu_1-\mu_2, \sigma_1^2/n_1+\sigma_2^2/n_2)$$

•D'où notre intervalle de confiance au niveau 1- α :

$$I_C(\mu_1-\mu_2)=\left[\bar{x}_1-\bar{x}_2-t_{n_1+n_2-2;\alpha/2}\hat{s}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}};\ \bar{x}_1-\bar{x}_2+t_{n_1+n_2-2;\alpha/2}\hat{s}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]$$

•Où $t_{\alpha/2}$ est lu dans la table de student à $n_1 + n_2 - 2$

2)- le cas de 2 populations quelconques:

a)- Si les deux variances σ_1^2 et σ_2^2 sont connues :

•On considère alors la V.A. pour $n_1, n_2 \geq 30$

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \approx \mathcal{N}(0,1)$$

$$I_C(\mu_1 - \mu_2) = \left[\bar{x}_1 - \bar{x}_2 - z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}; \; \bar{x}_1 - \bar{x}_2 + z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right]$$

- Donc, notre intervalle de confiance au niv. $1-\alpha$, de la différence des moyennes s'écrit :

$$I_C(\mu_1 - \mu_2) = \left[\bar{x}_1 - \bar{x}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}; \; \bar{x}_1 - \bar{x}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$$

V- Intervalle de Confiance pour la différence des proportions

•Soient P_1 et P_2 2 populations que l'on étudie selon la V.A. parente X suivant la loi $\mathcal{B}(p_1)$ sur P_1 et $\mathcal{B}(p_2)$ sur P_2.

•On veut estimer la différence des proportions $p_1 - p_2$. Alors dés que:

$$n_i \geq 30, \; n_i f_i \geq 5, \; n_i(1 - f_i) \geq 5 : i = 1,2$$

- <u>Si les variances sont inconnues:</u>
- On estime chaque σ_i^2 par S_i^2 et on considère alors la V.A. pour $n_1, n_2 \geq 50$

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \approx \mathcal{N}(0,1)$$

- Donc, notre intervalle de confiance au niv. $1-\alpha$, s'écrit :

- notre intervalle de confiance s'écrit au niveau $1-\alpha$:

$$I_C(p_1 - p_2) = \left[f_1 - f_2 - z_{\alpha/2}\sqrt{\frac{f_1(1-f_1)}{n_1} + \frac{f_2(1-f_2)}{n_2}}; \; f_1 - f_2 + z_{\alpha/2}\sqrt{\frac{f_1(1-f_1)}{n_1} + \frac{f_2(1-f_2)}{n_2}} \right]$$

VI- Intervalle de Confiance pour le Rapport des variances

•Les deux populations sont supposées être Normalement distribuées:

•Sur P_1 : X $\rightsquigarrow \mathcal{N}(\mu_1, \sigma_1^2)$

•Sur P_2 : X $\rightsquigarrow \mathcal{N}(\mu_2, \sigma_2^2)$

•On s'intéresse à l'estimation du rapport σ_1^2 / σ_2^2

• D'où notre intervalle de confiance pour le rapport de variance au niveau 1-α :

$$I_C\left(\frac{\sigma_1^2}{\sigma_2^2}\right) = \left(\frac{s_1^2}{s_2^2} F_{1-\alpha/2\ ;\ n_2-1, n_1-1};\ \frac{s_1^2}{s_2^2} F_{\alpha/2\ ;\ n_2-1, n_1-1}\right)$$

• **1)- Si les moyennes sont inconnues**

• On propose alors l'estimateur S_1^2 / S_2^2

• Or $Y_1 = \left[(n_1-1)S_1^2 / \sigma_1^2\right] \rightsquigarrow \chi^2_{n_1-1}$

$$Y_2 = \left[(n_2-1)S_2^2 / \sigma_2^2\right] \rightsquigarrow \chi^2_{n_2-1}$$

• D'où $\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \rightsquigarrow \mathcal{F}(n_1\text{-}1; n_2\text{-}1)$

• **Remarque :** Pour des raisons de lecture des tables statistiques de Fisher, on choisi d'estimer le rapport (σ_1^2/σ_2^2) si ($s_1^2 > s_2^2$); sinon on doit estimer le rapport (σ_2^2/σ_1^2).

$$1-\alpha = P\left(F_{1-\alpha/2;n_1-1,n_2-1} < \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} < F_{\alpha/2;n_1-1,n_2-1}\right)$$

$$= P\left(\frac{S_1^2}{S_2^2} F_{1-\alpha/2;n_2-1,n_1-1} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} F_{\alpha/2;n_2-1,n_1-1}\right)$$